

A corpus-based study of invariant tags in London English

Eivind Torgersen & Costas Gabrielatos
Lancaster University

This paper reports on the analysis of the use of a number of invariant tags in spoken London English, which formed part of the completed project *Analysis of spoken London English using corpus tools* (funded by the British Academy). The tags examined were: *innit*, *okay*, *right*, *yeah*, *you get me* and *you know*, as well as three semi-fixed expressions containing *you know*, which functioned as tags: *do you know what I mean*, *if you know what I mean* and *do you know what I'm saying*.

The study used the Linguistic Innovators Corpus (LIC), a 1.4 million word corpus comprising the transcribed and marked-up interview data from the Lancaster/Queen Mary ESRC-funded project, *Linguistic innovators: the English of adolescents in London* (Kerswill et al. 2008), as well as the Corpus of London Teenage English (COLT) (Stenström et al. 2002). The research methodology combined approaches and techniques from sociolinguistics and corpus linguistics. Five variables were examined, four of which had two values: age (young: 16-19 / old: over 70), sex (male/female), ethnicity (Anglo/non-Anglo) and place of residence (Hackney/Havering). The fifth variable was a self-assessed measure of the multi-ethnicity of the friendship networks that speakers belonged to, with scores ranging from 1 (all friends same ethnicity as self) to 5 (60%-80% of friends different ethnicity as self). The analysis took into account the relative frequency of use of each tag, as well as the proportion of speakers in each sociolinguistic group that used each tag.

The comparison of LIC and COLT revealed an increase in *yeah* and, in particular, *innit*, and a dramatic increase in *you get me*, but a decrease in the relative frequencies of *right* and *okay*. The analysis of LIC showed that all the innovative tags, such as *innit* and *you get me*, were clearly a feature of young people's speech. In addition, the most innovative tag, *you get me*, was by far most frequent in Hackney (inner London), and the highest frequency was observed among the non-Anglo speakers. The ethnic minority speakers, and male speakers in general, are the most innovative tag users, particularly of *innit* and *you get me*, but the ethnic minority speakers also had high frequencies of *yeah*, *okay* and *right*, and they were therefore the highest users of tags overall. Overall, there is a difference in tag usage between inner and outer London: the more innovative tags are more frequent in inner London and the more traditional ones in outer London. The innovative tags *you get me* and *innit* were most frequent, and were used by a larger proportion of speakers, among male, non-Anglo, Hackney residents.

The results indicate that young people, ethnic minorities, an urban environment, and dialect contact are of great importance in language change, a fact that can feed into an exploratory model of language variation and change.

Corpus Linguistics 2009
University of Liverpool
23 July 2009

A corpus-based study of invariant tags in London English

Eivind Torgersen & Costas Gabrielatos

Lancaster University

e.torgersen@lancaster.ac.uk; c.gabrielatos@lancaster.ac.uk



Why study London English?

- London as the centre of linguistic innovation in British English
 - Diffusion of linguistic features from inner to outer London and beyond
- London as a multicultural city
 - High level of dialect and language contact

1970-1979

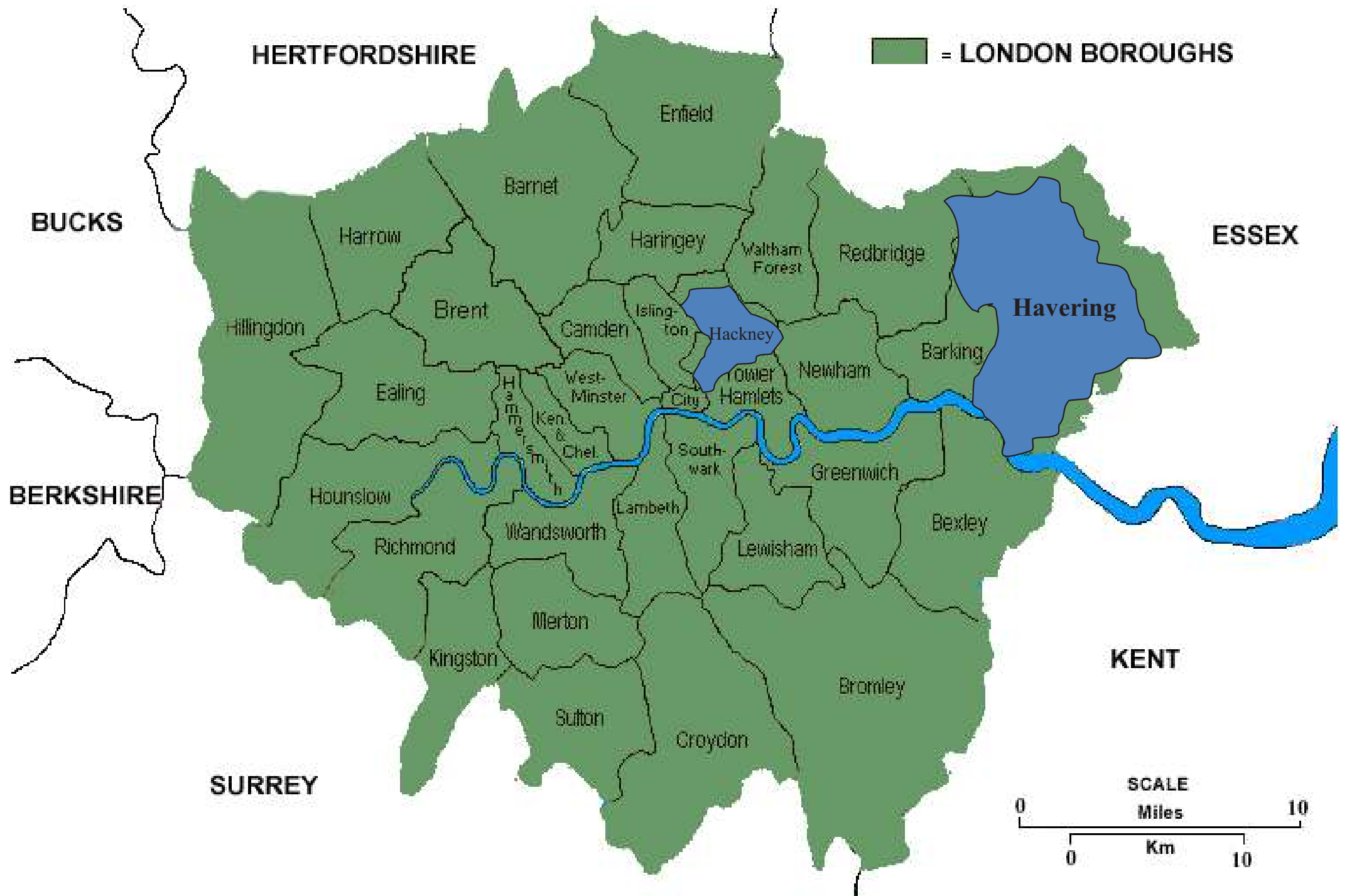
1980-1989 (1980-1989) (1980-1989)

1990-1999

2000-2009 (2000-2009) (2000-2009)

2010-2019 (2010-2019) (2010-2019)

E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL



Invariant tags

- Part of a tag question
- Anchor and tag
 - Canonical tag
 - *It's cold, isn't it?*
 - *They're late, aren't they?*
 - Invariant tag
 - *It's cold, innit?*
 - *They're late, innit?*

Why study invariant tags?

- Frequent in spoken language
- Frequent in young people's speech
- Young people are linguistic innovators
- An innovative tag user = A linguistic innovator?

We examined:

- Simple invariant tags
 - *innit, okay, right, yeah*
- Multi-word invariant tags: (elliptical) clauses
 - *You get me*
 - *You know*
 - *(Do)/(If) (you) know what I mean*
 - *(Do) (you) know what I'm saying*

Some examples

- but he's been here for her **innit**?
- they just ain't got nothing **innit**?
- she's coming up for sixty now **yeah**?
- no the thing **right**? I I didn't mind **right**? but the thing that pissed me off is the that she brung the fight into the house
- he makes you laugh but he's just annoying **if you know what I mean**? but he's just he will come behind me once **yeah**? he come behind me and he's got my hood on my jacket and stuck it over my head

Linguistic Innovators Corpus (LIC)

- 2005/2008
- Sociolinguistic interviews
- 1.3 million words
- Speaker data:
 - Sex
 - Age: young (16-19), old (60+)
 - Residence: Hackney (inner London), Havering (outer London)
 - Ethnicity: Anglo, Non-Anglo
 - Friendship network score (% of different ethnicity friends):
1 (0%) – 5 (60-80%)
 - All speakers are working class

Corpus of London Teenage Language (COLT)

- 1993
- 500,000 words
- Self recordings
- Speaker data:
 - Sex
 - Age
 - Residence: Hackney, Tower Hamlets, Camden (inner London); Barnet (outer London), Hertfordshire
 - Social class
- COLT also contains speech by middle-class recruits and unknown speakers.

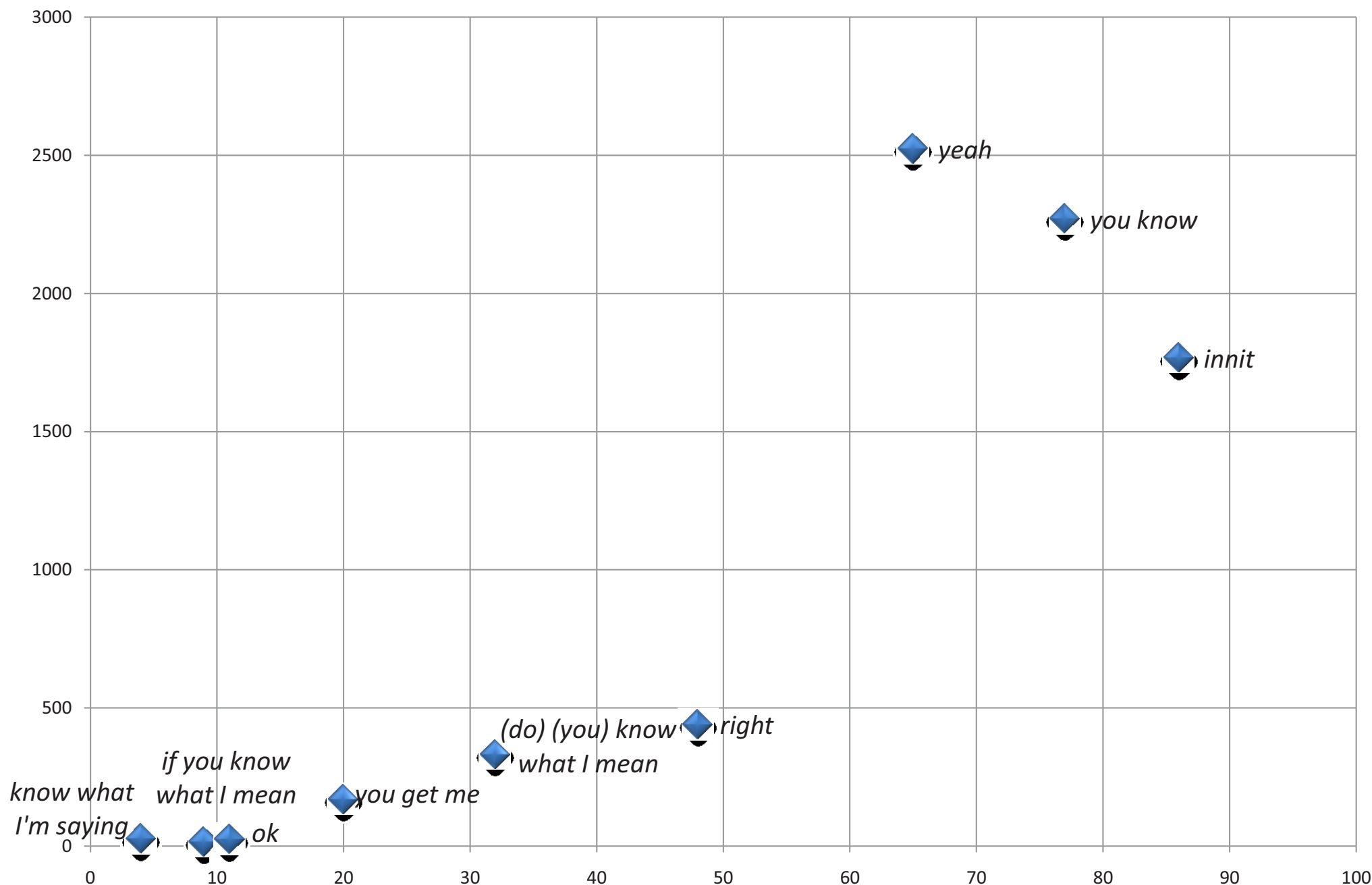
Methodology

- Automatic extraction of concordances of candidate word-forms – with mark up (by Sebastian Hoffmann)
- Manual identification of tags:
 - Transcript (concordance analysis): co-text
 - Recordings: phonological features
- Tabulation of tag instances – with sociolinguistic information

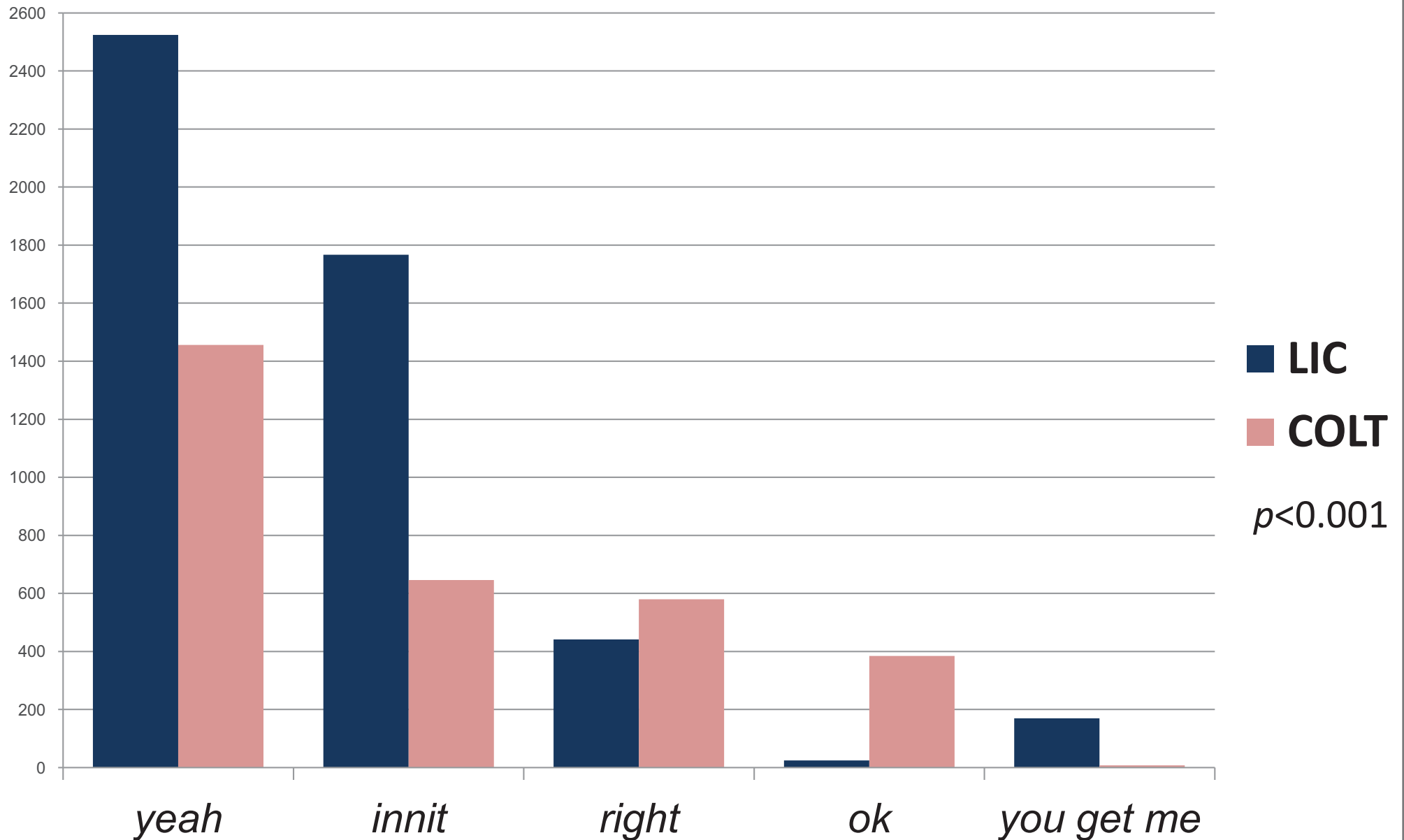
Metrics

- **Frequency**
 - Occurrences per million words
- **Spread**
 - Ratio (%) of tag users to speakers

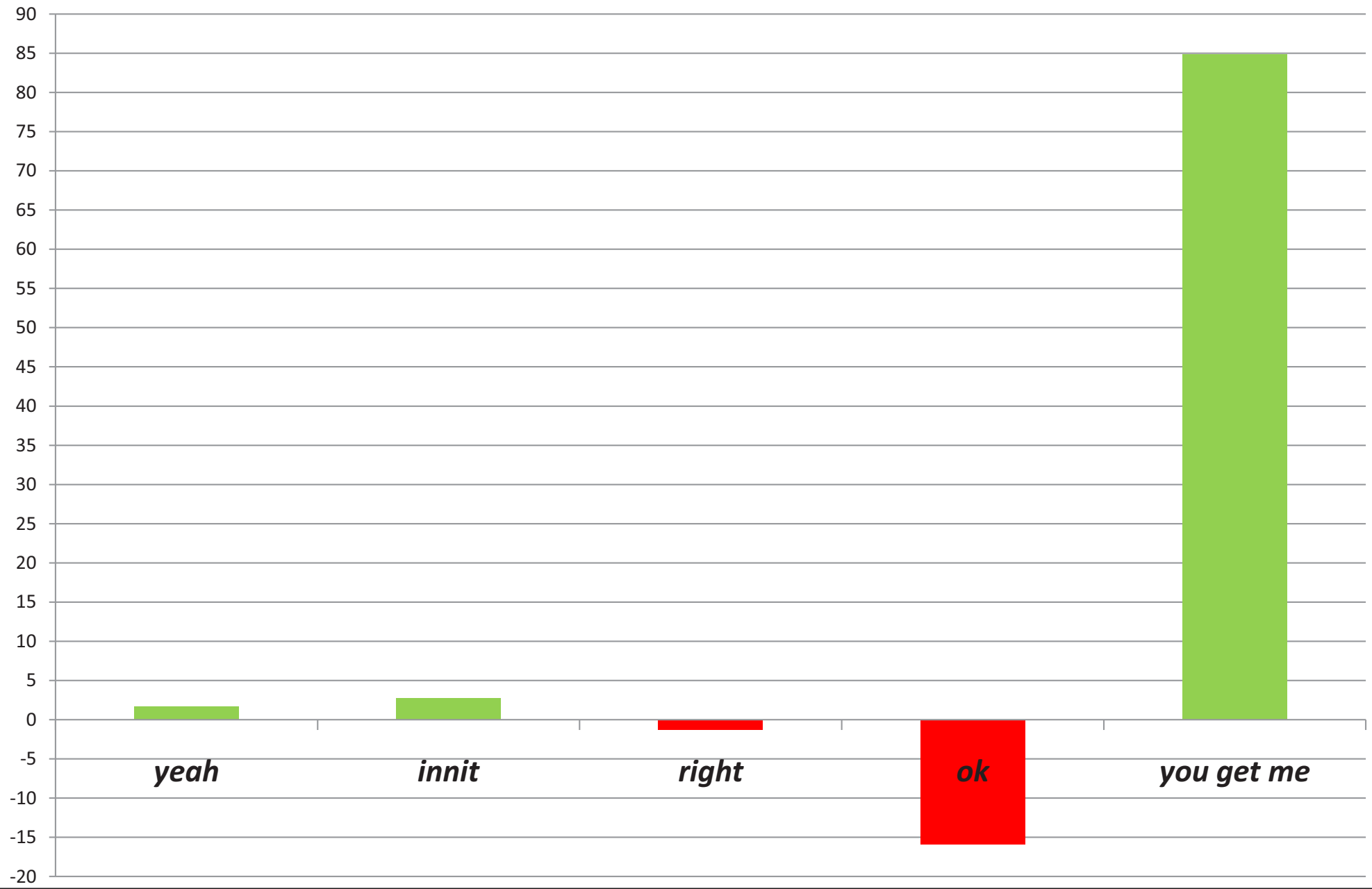
Young LIC speakers: frequency and spread



Young speakers in LIC and COLT: Frequency comparison



Young speakers in LIC and COLT: Difference ratio LIC/COLT



Looking at tags:
Use by sociolinguistic variables

- **Bold:** both frequency and spread differences are statistically significant.
- Normal: only frequency differences are statistically significant.
- '-----': both frequency and spread are comparable.

	Age	Sex	Ethnicity	Residence
<i>yeah</i>	Young	-----	Non-Anglo	-----
<i>you know</i>	Old	Male	Anglo	-----
<i>innit</i>	Young	Male	Non-Anglo	-----
<i>ok</i>	-----	Female	Non-Anglo	-----
<i>right</i>	Young	Female	Non-Anglo	Hackney
<i>(do) (you) know what I mean</i>	Young	Female	Anglo	Havering
<i>if you know what I mean</i>	Young	-----	-----	Havering
<i>(do) you know what I'm saying</i>	Young	Female	-----	Havering
<i>you get me</i>	Young	-----	Non-Anglo	Hackney

High-frequency tags

	Age	Sex	Ethnicity	Residence
<i>innit</i>	Young	Male	Non-Anglo	-----
<i>yeah</i>	Young	-----	Non-Anglo	-----
<i>you know</i>	Old	Male	Anglo	-----
<i>ok</i>	-----	Female	Non-Anglo	-----
<i>right</i>	Young	Female	Non-Anglo	Hackney
<i>(do) (you) know what I mean</i>	Young	Female	Anglo	Havering
<i>if you know what I mean</i>	Young	-----	-----	Havering
<i>(do) you know what I'm saying</i>	Young	Female	-----	Havering
<i>you get me</i>	Young	-----	Non-Anglo	Hackney

Low-frequency simple tags

	Age	Sex	Ethnicity	Residence
<i>innit</i>	Young	Male	Non-Anglo	-----
<i>yeah</i>	Young	-----	Non-Anglo	-----
<i>you know</i>	Old	Male	Anglo	-----
<i>ok</i>	-----	Female	Non-Anglo	-----
<i>right</i>	Young	Female	Non-Anglo	Hackney
<i>(do) (you) know what I mean</i>	Young	Female	Anglo	Havering
<i>if you know what I mean</i>	Young	-----	-----	Havering
<i>(do) you know what I'm saying</i>	Young	Female	-----	Havering
<i>you get me</i>	Young	-----	Non-Anglo	Hackney

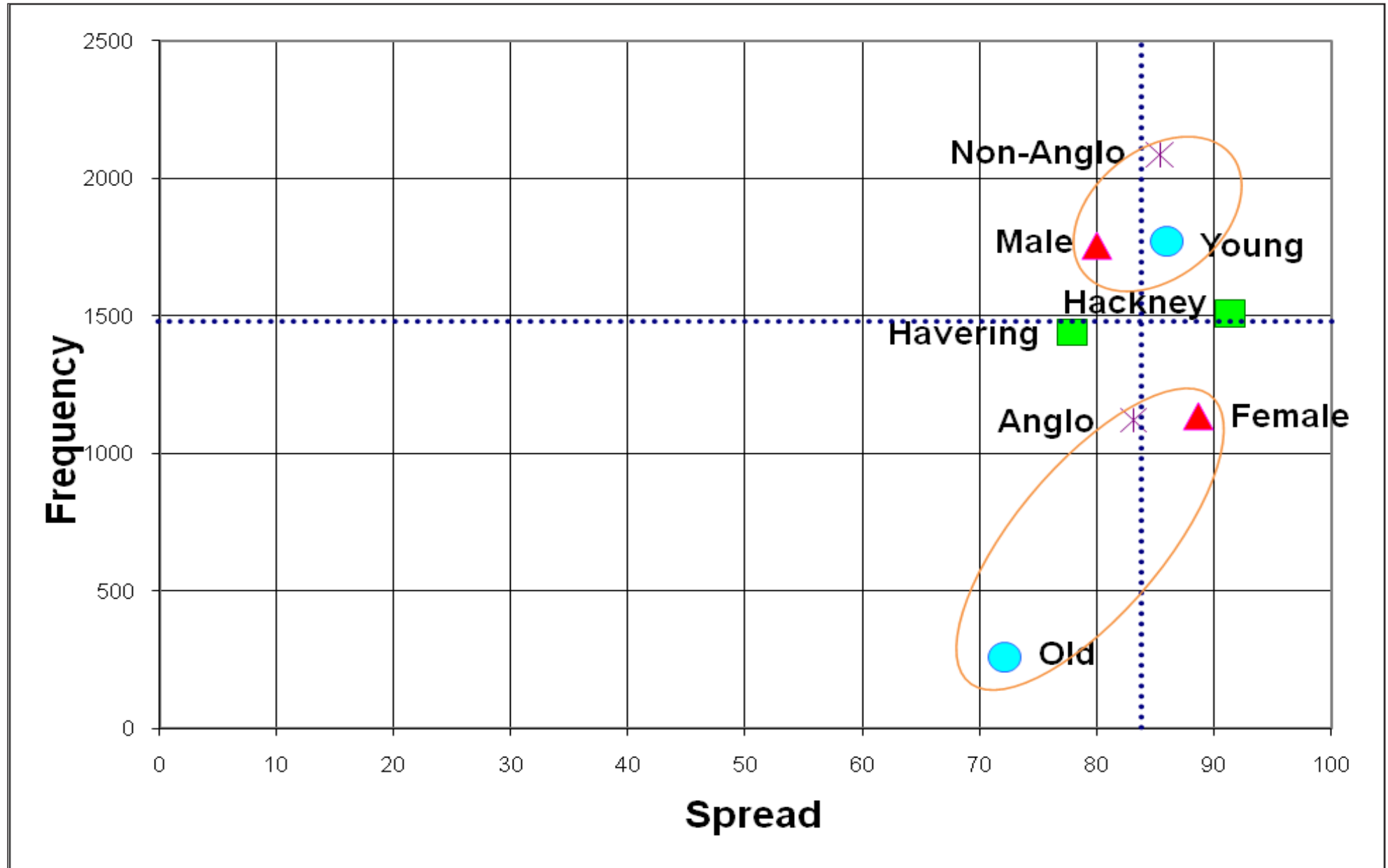
Low-frequency multi-word tags

	Age	Sex	Ethnicity	Residence
<i>innit</i>	Young	Male	Non-Anglo	-----
<i>yeah</i>	Young	-----	Non-Anglo	-----
<i>you know</i>	Old	Male	Anglo	-----
<i>ok</i>	-----	Female	Non-Anglo	-----
<i>right</i>	Young	Female	Non-Anglo	Hackney
<i>(do) (you) know what I mean</i>	Young	Female	Anglo	Havering
<i>if you know what I mean</i>	Young	-----	-----	Havering
<i>(do) you know what I'm saying</i>	Young	Female	-----	Havering
<i>you get me</i>	Young	-----	Non-Anglo	Hackney

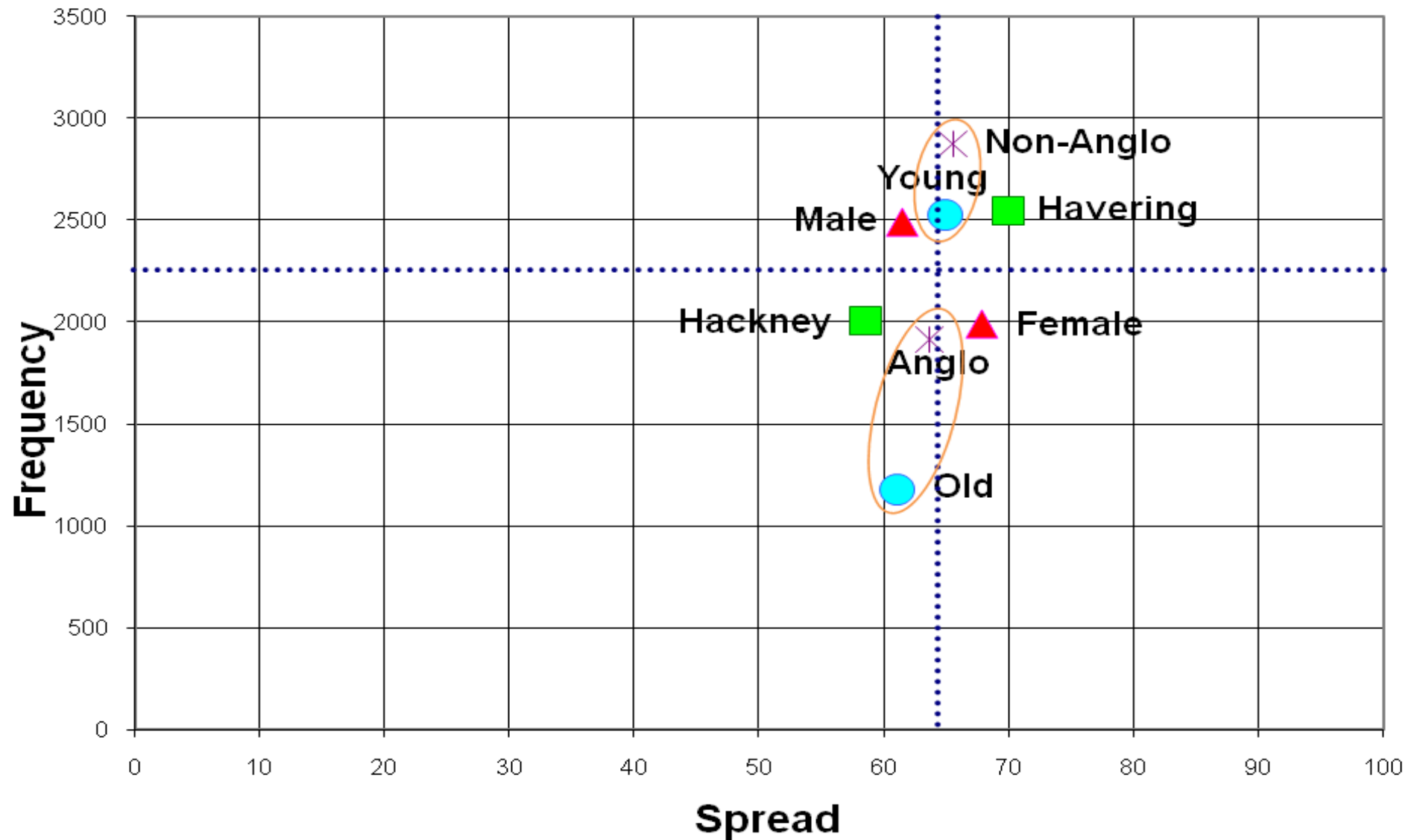
Emerging tag

	Age	Sex	Ethnicity	Residence
<i>innit</i>	Young	Male	Non-Anglo	-----
<i>yeah</i>	Young	-----	Non-Anglo	-----
<i>you know</i>	Old	Male	Anglo	-----
<i>ok</i>	-----	Female	Non-Anglo	-----
<i>right</i>	Young	Female	Non-Anglo	Hackney
<i>(do) (you) know what I mean</i>	Young	Female	Anglo	Havering
<i>if you know what I mean</i>	Young	-----	-----	Havering
<i>(do) you know what I'm saying</i>	Young	Female	-----	Havering
<i>you get me</i>	Young	-----	Non-Anglo	Hackney

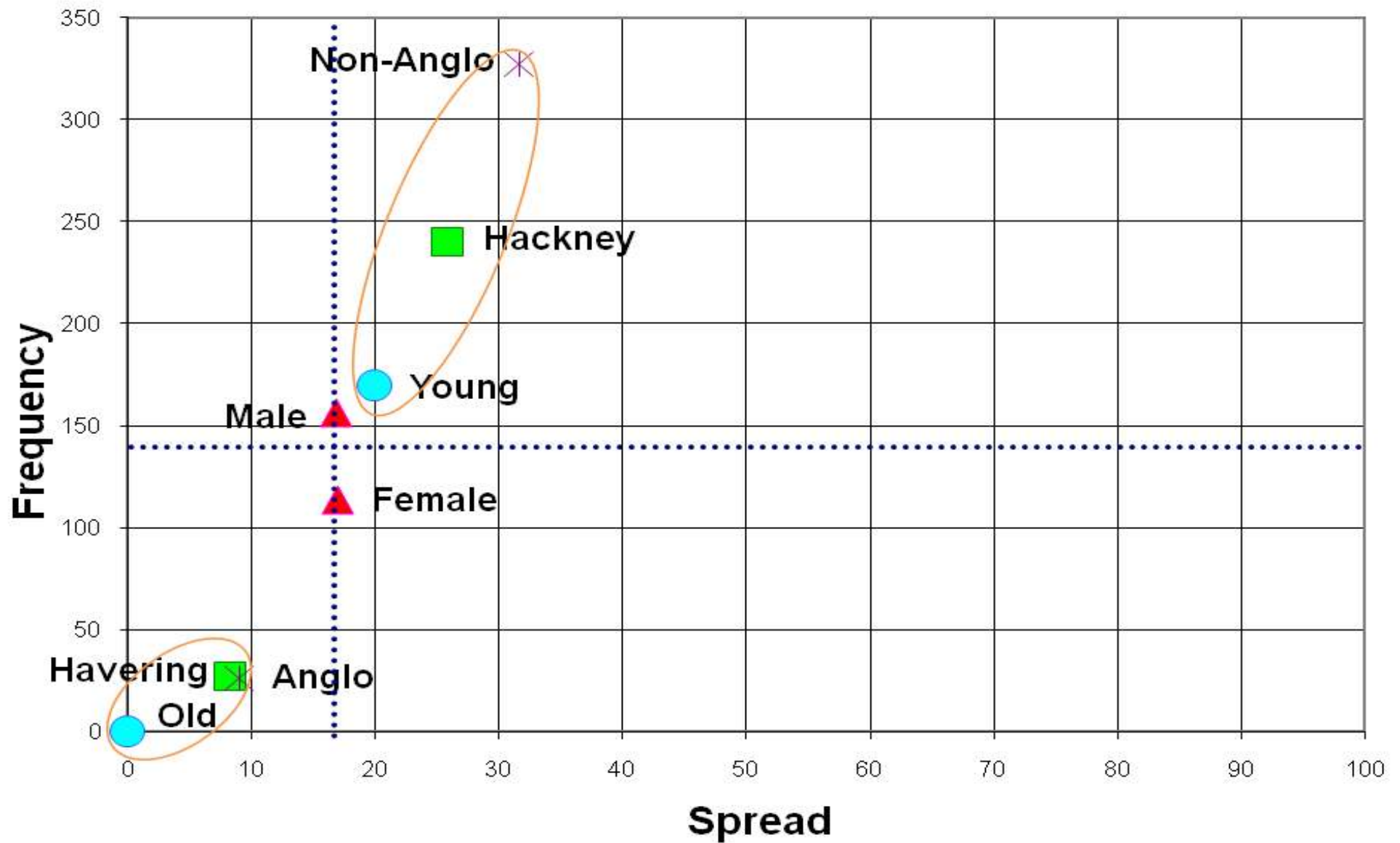
innit



yeah



you get me



Looking at speakers: Tag usage in sociolinguistic groups

Bold indicates that
both frequency and spread differences
are statistically significant

Age

	Young	Old	No clear preference
Freq. + Spread	<ul style="list-style-type: none"> <i>you get me</i> 	-----	<ul style="list-style-type: none"> <i>ok</i>
Freq. only	<ul style="list-style-type: none"> <i>innit</i> <i>right</i> <i>(do) (you) know what I mean</i> <i>if you know what I mean</i> <i>(do) you know what I'm saying</i> <i>yeah</i> 	<ul style="list-style-type: none"> <i>you know</i> 	

- Vast majority of tags more frequently used by young speakers.

Sex

	Male	Female	No clear preference
Freq. + Spread	-----	<ul style="list-style-type: none"> • <i>ok</i> 	<ul style="list-style-type: none"> • <i>you get me</i> • <i>if you know what I mean</i> • <i>yeah</i>
Freq. only	<ul style="list-style-type: none"> • <i>innit</i> • <i>you know</i> 	<ul style="list-style-type: none"> • <i>right</i> • <i>(do) (you) know what I mean</i> • <i>(do) you know what I'm saying</i> 	

- Female speakers show preference for a larger number of tags.
- Male speakers show preference for the high frequency tags.

Ethnicity

	Anglo	non-Anglo	No clear preference
Freq. + Spread	-----	<ul style="list-style-type: none"> <i>you get me</i> 	<ul style="list-style-type: none"> <i>if you know what I mean</i> <i>(do) you know what I'm saying</i>
Freq. only	<ul style="list-style-type: none"> <i>you know</i> <i>(do) (you) know what I mean</i> 	<ul style="list-style-type: none"> <i>innit</i> <i>ok</i> <i>right</i> <i>yeah</i> 	

- Non-Anglo speakers ...
 - show preference for a larger number of tags.
 - have a significantly higher frequency of all simple tags
 - show significantly higher frequency and spread for the innovative tag *you get me*.

Inner and outer city

	Hackney	Havering	No clear preference
Freq. + Spread	<ul style="list-style-type: none"> <i>you get me</i> 	-----	<ul style="list-style-type: none"> <i>innit</i> <i>ok</i> <i>yeah</i> <i>you know</i>
Freq. only	<ul style="list-style-type: none"> <i>right</i> 	<ul style="list-style-type: none"> <i>(do) (you) know what I mean</i> <i>if you know what I mean</i> <i>(do) you know what I'm saying</i> 	

- *you get me* and *right* characteristic of Hackney speakers
- The multi-word '*you know*' tags characteristic of Havering speakers.
- Comparable preference for most simple tags.

You get me: Use and friendship network

Comparison 1:

Average network scores

users vs. non-users

- Users have a 30% higher average network score than non-users

Comparison 2:

Number of (non-)users

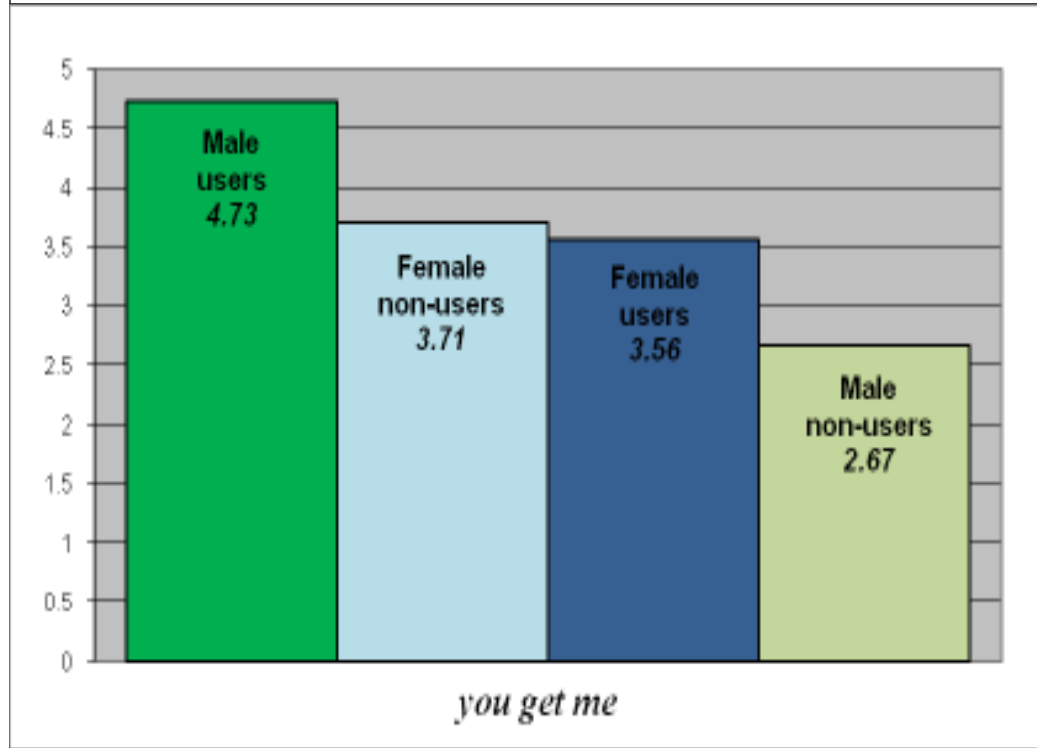
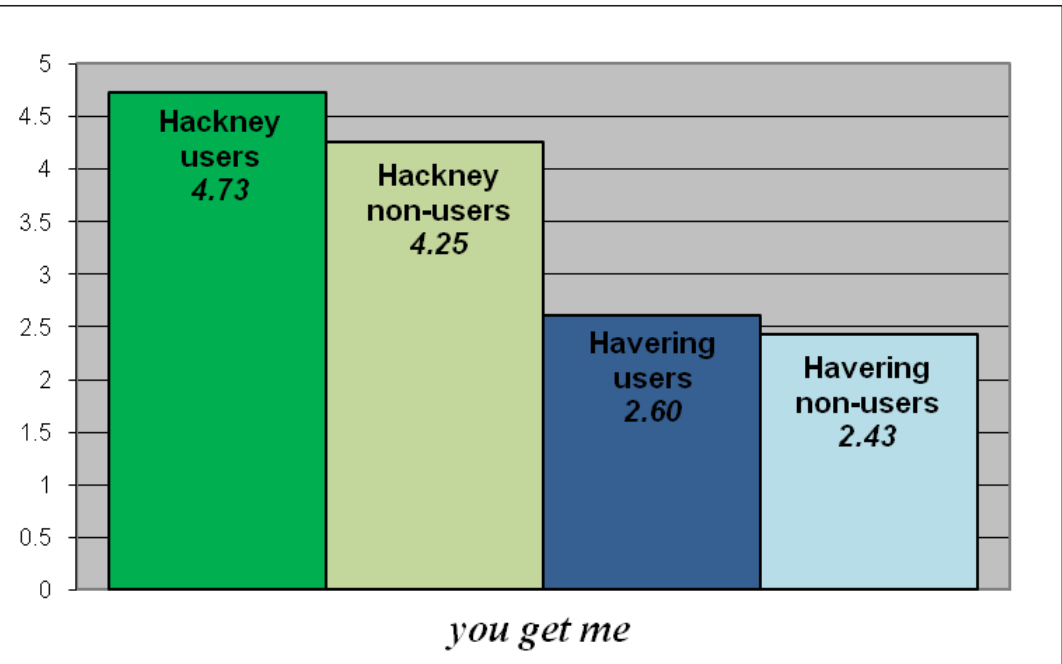
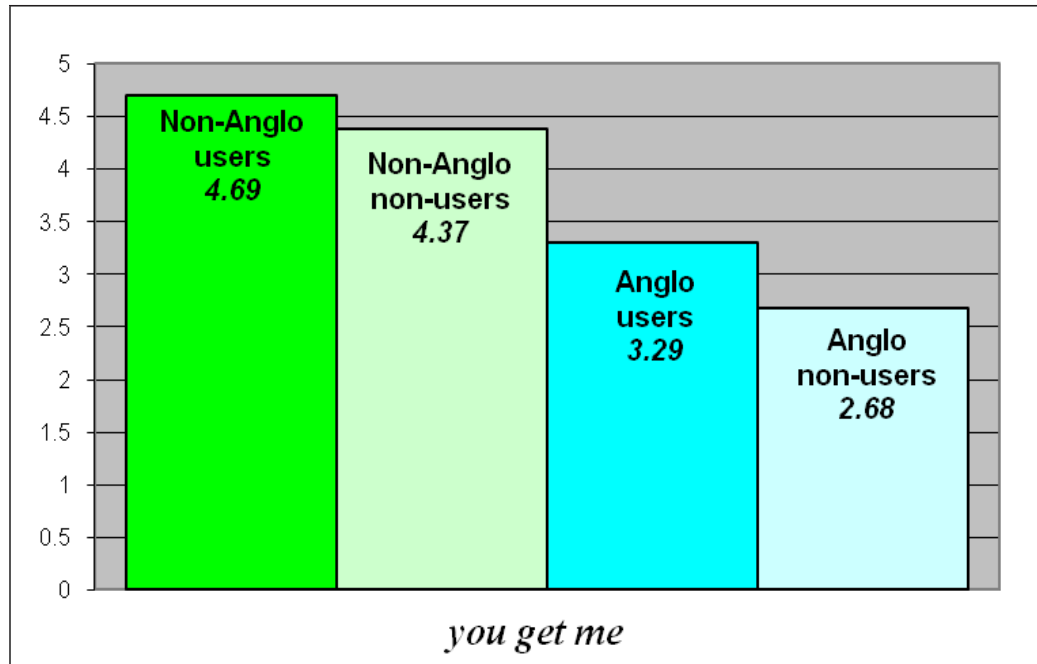
high-score users vs. low-score users

high-score non-users vs. low-score non-users

- Users: high-score six times more than low-score.
- Non-users: comparable numbers

- Users of *you get me* can be expected to belong to multi-ethnic friendship networks.
 - No safe predictions can be made regarding the friendship networks of non-users.
- Tag still emerging

You get me: Network score and use



On its own, network score does not seem to be a dependable predictor of use of *you get me*.

Innovative tags and linguistic innovation

- Established tags, irrespective of whether they are becoming more or less frequent, have a less marked ethnic distribution.
- New (innovative) tags, like *you get me*, are currently used significantly more frequently within the multi-ethnic networks in which they have probably first emerged.

Tags as part of the bigger picture

- Innovation in the invariant tag system is in line with previous findings on innovation in phonology and grammar.
- Inner-city, non-Anglo males are in the lead.
- The frequency of new tags is highest in high-density multi-cultural friendship groups.
- Our current research focuses on the functional distributions of tags